Problem 1

Problem 2

Problem 3

$$\frac{\partial}{\partial x}(\tanh(x)) = \frac{\partial}{\partial x}(\frac{e^x - e^{-x}}{e^x + e^{-x}})$$
$$= \frac{(e^x + e^{-x})(e^x + e^{-x}) - (e^x - e^{-x})(e^x - e^{-x})}{(e^x + e^{-x})^2}$$
$$= 1 - \tanh^2(x)$$

It substantially simplifies the calculations during backpropagation when computing the local gradient at an intermediate node and when passing the gradient to the lower layers.

Problem 4

Problem 5

Problem 6

We would stop training approximately at iteration 50, because while the training error keeps decreasing we have an increase in the validation error, which is a sign of overfitting. After we take the weights of the network at this iteration we test it one final time on the test set. Note that we do not get the optimal test error as we would if we would continue with the training until iteration 80.

Problem 7

Problem 8

$$\frac{e^{x_i}}{\sum_{i=1}^{N} e^{x_i}} = \frac{e^a}{e^a} \frac{e^{x_i}}{\sum_{i=1}^{N} e^{x_i}}$$
$$= \frac{e^{x_i - a}}{\sum_{i=1}^{N} e^{x_i - a}}$$

Problem 9